

# Perfect Sampling with Unitary Tensor Networks

Andrew J. Ferris<sup>1,2</sup> and Guifre Vidal<sup>1,3</sup>

<sup>1</sup>*The University of Queensland, School of Mathematics and Physics, Queensland 4072, Australia*

<sup>2</sup>*Département de Physique, Université de Sherbrooke, Québec, J1K 2R1, Canada*

<sup>3</sup>*Perimeter Institute for Theoretical Physics, Waterloo, Ontario, N2L 2Y5, Canada*

(Dated: April 11, 2012)

Tensor network states are powerful variational ansätze for many-body ground states of quantum lattice models. The use of Monte Carlo sampling techniques in tensor network approaches significantly reduces the cost of tensor contractions, potentially leading to a substantial increase in computational efficiency. Previous proposals are based on a Markov chain Monte Carlo scheme generated by locally updating configurations and, as such, must deal with equilibration and autocorrelation times, which result in a reduction of efficiency. Here we propose perfect sampling schemes, with vanishing equilibration and autocorrelation times, for unitary tensor networks – namely tensor networks based on efficiently contractible, unitary quantum circuits, such as unitary versions of the matrix product state (MPS) and tree tensor network (TTN), and the multi-scale entanglement renormalization ansatz (MERA). Configurations are directly sampled according to their probabilities in the wave-function, without resorting to a Markov chain process. We consider both complete sampling, involving all the relevant sites of the system, as well as incomplete sampling, which only involves a subset of those sites, and which can result in a dramatic (basis-dependent) reduction of sampling error.

PACS numbers: 05.10.–a, 02.50.Ng, 03.67.–a, 74.40.Kb

## I. INTRODUCTION

To the computational physicist interested in one-dimensional quantum lattice models, the density matrix renormalization group (DMRG)<sup>1,2</sup> is a dream come true. It provides an essentially unbiased, extremely accurate variational approach to ground state properties of a large class of local Hamiltonians in one dimensional lattices. DMRG operates by approximating the ground state of the system with a matrix product state (MPS)<sup>3–6</sup>, which is a simple tensor network with tensors connected according to a one-dimensional array. In recent years, the success and broad applicability of DMRG has been understood to follow from (i) the existence of a characteristic, universal pattern of entanglement common to most ground states in one spatial dimension; and (ii) the ability of the MPS to reproduce this universal pattern of entanglement, thanks to having its tensors connected into a one-dimensional geometry.

The above insight has since then guided the development of new tensor network approaches that aim to repeat, in other geometries or physical regimes of interest, the unprecedented success of DMRG<sup>1,2,7,8</sup> in one dimension. The recipe is quite simple: first, identify a pattern of entanglement common to a large class of ground states; then, connect tensors so that they can reproduce this pattern, and use the resulting tensor network as a variational ansatz. In this way the multi-scale, layered pattern of entanglement observed in ground states near a continuous quantum phase transition motivated the proposal of the multi-scale entanglement renormalization ansatz (MERA)<sup>9,10</sup> to address quantum critical phenomena. Similarly, the characteristic spatial pattern of entanglement in the ground states in two and higher dimensions motivated higher-dimensional generalizations of both the MPS (known as projected entangled-pair states, PEPS<sup>11–17</sup>) and the MERA<sup>18–21</sup>.

The cost of simulating a lattice of  $L$  sites with any of the above tensor networks is roughly proportional to  $L$ , which un-

derlies the efficiency of the approaches<sup>22</sup>. Importantly, however, this cost also grows as  $O(\chi^p)$ , that is as a power  $p$  of the dimension  $\chi$  of the indices connecting the tensors into a network. On the one hand, this bond dimension  $\chi$  determines the size of the tensors and therefore the number of variational parameters contained in the tensor network ansatz. On the other,  $\chi$  is also a measure of how much entanglement the tensor network can carry. It then follows that the cost of simulations increases with the amount of entanglement in the ground state of the system. Entanglement is indeed the key factor limiting the range of applicability of tensor network approaches.

More specifically, for an MPS, a small power  $p$ , namely  $p_{\text{MPS}} = 3$ , implies that very large values of  $\chi$  (of up to a few thousands) can be considered even with a high-end desktop computer. Correspondingly, DMRG can address one-dimensional systems with robustly entangled ground states. In contrast, the cost of two dimensional simulations with PEPS and MERA scales with a much larger power  $p$  of  $\chi$ , e.g.  $p_{\text{PEPS}} = 12$  in Ref. 14 and  $p_{\text{MERA}} = 16$  in Ref. 21, and this considerably reduces the affordable values of  $\chi$ . In other words, PEPS and MERA calculations have so far been restricted to systems with relatively small amounts of ground state entanglement. A major present challenge for these approaches is to obtain more efficient tensor contraction schemes that could lower their cost.

A possible route to reducing the scaling of computational cost with  $\chi$  in tensor network algorithms is by using Monte Carlo sampling techniques, as proposed in Refs. 23–25. As reviewed in the next section, the cost of manipulating the tensor network (for a single sample) is reduced to  $O(\chi^q)$ , where  $q$  is significantly smaller than  $p$  (typically of the order of  $p/2$ ). The proposals in Refs. 23,24 are best suited for tensor networks, such as MPS and PEPS, where the coefficients in the tensors are unconstrained. However, in the MERA, as well as in other unitary tensor networks such as unitary versions of MPS (uMPS) and of tree tensor network<sup>26,27</sup> (uTTN), tensors

are subject to unitary constraints.

The purpose of this paper is to address the use of Monte Carlo sampling in the context of unitary tensor networks, including uMPS, uTTN and MERA. [Notice that this excludes tensor networks such as a periodic MPS or PEPS, which cannot be generically re-expressed as a unitary tensor network]. An important difference with respect to Refs. 23,24 is that in a unitary tensor network, sampling is performed on an effective lattice corresponding to the past causal cone of the local operator whose expectation value is being computed. This means that sampling typically occurs over some reduced number of sites (less than the system size  $L$ ). A second difference is that in unitary tensor networks there is no need to use a Markov chain Monte Carlo scheme. Indeed, our main result is the proposal and benchmark of perfect sampling schemes for unitary tensor networks, by means of which one can obtain completely uncorrelated samples directly according to the correct probability. Therefore, one can sample without incurring additional computational costs due to equilibration and autocorrelations times. This is particularly of interest near a quantum phase transition, where equilibration and autocorrelation times diverge with system size  $L$ . We consider both complete (perfect) sampling and incomplete (perfect) sampling schemes. In the former, the indices for all sites of the effective lattice are sampled. In the latter, only the indices of a subset of sites is sampled, while the indices of the rest of sites are contracted exactly, with an insignificant or minor increase of computational cost as far as the scaling  $O(\chi^q)$  is concerned. Importantly, the statistical variance (due to sampling) of an expectation value obtained with incomplete sampling can decrease dramatically with a proper choice of sampling basis, as illustrated in Fig. 9 with a drop of  $10^{-7}$  in error.

The paper is organized in sections as follows. First, in section II we briefly review the use of Monte Carlo sampling techniques to evaluate the expectation value of local operators in context of tensor networks, and introduce the notions of complete and incomplete sampling. Then in section III we explain how the proposals of Refs. 23,24 can be adapted to the case of a unitary tensor network by sampling within the past causal cone of the local operator. In section IV we propose a complete perfect sampling scheme for unitary tensor networks. Its performance is demonstrated for a uMPS with the quantum Ising chain at criticality. In section V we then present an incomplete perfect sampling scheme. We discuss computational costs in section VI. The conclusions in Section VII and an Appendix analyzing the variance in different schemes close the paper.

We emphasize that this paper is only concerned with the evaluation of local expectation values from a unitary tensor network. That is, here we assume that the unitary tensor network has already been optimized and focus on how to extract information from it. The optimization of unitary tensor networks using variational Monte Carlo is discussed in Ref. 29.

## II. BACKGROUND MATERIAL: SAMPLING IN TENSOR NETWORK ALGORITHMS

Let us start by introducing our notation and by reviewing some basic concepts.

### A. Exact contraction versus sampling

Let  $\mathcal{L}$  be a lattice made of  $L$  sites, with vector space  $\mathbb{V}_{\mathcal{L}} \equiv \otimes_{i=1}^L \mathbb{V}$ , where  $\mathbb{V}$  is the  $d$ -dimensional vector space of one site. Let  $|\Psi\rangle \in \mathbb{V}_{\mathcal{L}}$  denote the wave-function encoded in the tensor network and let  $\hat{A}$  be a local operator on  $\mathbb{V}_{\mathcal{L}}$ . An important task in tensor network algorithms is to compute the expectation value  $\langle \Psi | \hat{A} | \Psi \rangle$ , which can be expressed as

$$\langle \Psi | \hat{A} | \Psi \rangle = \sum_{\mathbf{s} \in \mathcal{S}} \langle \Psi | \mathbf{s} \rangle \langle \mathbf{s} | \hat{A} | \Psi \rangle, \quad (1)$$

where  $|\mathbf{s}\rangle \equiv |s_1\rangle \otimes |s_2\rangle \otimes \cdots \otimes |s_L\rangle$  denotes a product state of the  $L$  sites of the lattice, with  $s_i = 1, 2, \dots, d$  labelling the elements of an orthonormal basis  $\{|s_i\rangle\}$  on site  $i$ ,  $i = 1, 2, \dots, L$ . Here,  $\mathcal{S}$  is the set of all  $d^L$  possible *configurations*  $\mathbf{s} = (s_1, s_2, \dots, s_L)$  of the system. The expectation value of Eq. (1) can be obtained exactly by contracting the corresponding tensor network. However, a large computational cost motivates the search for an alternative approach based on sampling.

In preparation for an approximate evaluation of the expectation value  $\langle \Psi | \hat{A} | \Psi \rangle$ , let us first introduce the probability  $Q(\mathbf{s}) \equiv |\langle \mathbf{s} | \Psi \rangle|^2$  of projecting state  $|\Psi\rangle$  into the product state  $|\mathbf{s}\rangle$ , and the estimator  $A(\mathbf{s}) \equiv \langle \mathbf{s} | \hat{A} | \Psi \rangle / \langle \mathbf{s} | \Psi \rangle$ , and rewrite Eq. (1) as

$$\langle \Psi | \hat{A} | \Psi \rangle = \sum_{\mathbf{s} \in \mathcal{S}} Q(\mathbf{s}) A(\mathbf{s}). \quad (2)$$

This expression emphasizes that  $\langle \Psi | \hat{A} | \Psi \rangle$  can be regarded as a probabilistic average of estimator  $A(\mathbf{s})$  according to the probabilities  $Q(\mathbf{s})$ , where  $Q(\mathbf{s}) \geq 0$ ,  $\sum_{\mathbf{s} \in \mathcal{S}} Q(\mathbf{s}) = 1$ .

Let us replace the sum over the set  $\mathcal{S}$  of all  $|\mathcal{S}| = d^L$  configurations  $\mathbf{s}$  with a sum over some subset  $\tilde{\mathcal{S}} \subseteq \mathcal{S}$  containing  $N \equiv |\tilde{\mathcal{S}}|$  configurations  $\mathbf{s}$ , where  $N < d^L$ , that is

$$\langle \Psi | \hat{A} | \Psi \rangle \approx \frac{1}{Z} \sum_{\mathbf{s} \in \tilde{\mathcal{S}}} Q(\mathbf{s}) A(\mathbf{s}), \quad (3)$$

where  $Z \equiv \sum_{\mathbf{s} \in \tilde{\mathcal{S}}} Q(\mathbf{s})$  is a normalization factor. Eq. (3) states that an approximate evaluation of  $\langle \Psi | \hat{A} | \Psi \rangle$  is obtained by considering a probabilistic sum over  $N$  configurations  $\mathbf{s}$ . If the  $N$  configurations in  $\tilde{\mathcal{S}}$  have been randomly chosen from  $\mathcal{S}$  according to the probability  $Q(\mathbf{s})$ , then importance sampling allow us to replace the previous expression with

$$\langle \Psi | \hat{A} | \Psi \rangle \approx \frac{1}{N} \sum_{\mathbf{s} \in \tilde{\mathcal{S}}} A(\mathbf{s}). \quad (4)$$

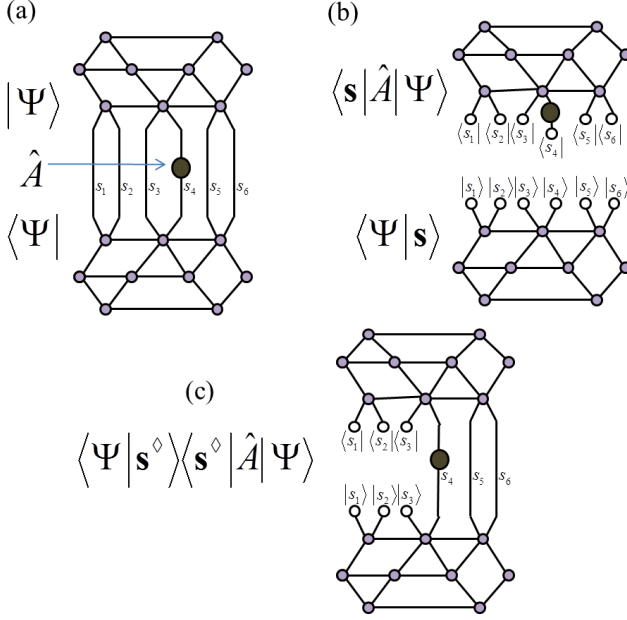


FIG. 1: (Color online) Contraction of a tensor network. (a) Tensor network corresponding to the expectation value  $\langle \Psi | \hat{A} | \Psi \rangle$ , with a sum over (or exact contraction of) indices  $s_1, s_2, \dots, s_6$  (exact contraction). Contracting this tensor network has a cost that scales as  $O(\chi^p)$  with the bond index  $\chi$ , for some power  $p$ . (b) Tensor networks corresponding to  $\langle \Psi | s \rangle \langle s | \hat{A} | \Psi \rangle$  for a given configuration  $s$ , corresponding to a single sample. The cost of contracting these two networks scales as  $O(\chi^q)$  with the bond index  $\chi$ , where power  $q$  is smaller than power  $p$ . (c) Tensor network corresponding to  $\langle \Psi | s^\diamond \rangle \langle s^\diamond | \hat{A} | \Psi \rangle$  for a given incomplete configuration  $s^\diamond \equiv (s_1, s_2, s_3)$  (these three indices are being sampled), where in addition there is a sum over (or exact contraction of) indices  $s_4, s_5$  and  $s_6$ . The cost of contracting this tensor network scales as  $O(\chi^{q'})$ , with  $q'$  somewhere between  $q$  and  $p$ .

Equation (4) estimates  $\langle \Psi | \hat{A} | \Psi \rangle$  by means of  $N$  independent samples of a random variable  $(A(s), Q(s))$ . By construction, the mean  $\bar{A}$  of this random variable,

$$\bar{A} \equiv \sum_{s \in \mathcal{S}} Q(s) A(s), \quad (5)$$

is given by the expectation value  $\langle \Psi | \hat{A} | \Psi \rangle$  of operator  $\hat{A}$ , see Eq. (2). Notice that, in addition, its variance  $\sigma_A^2$ , defined by

$$\sigma_A^2 \equiv \sum_s Q(s) |A(s) - \bar{A}|^2 \quad (6)$$

$$= \sum_s Q(s) |A(s)|^2 - |\bar{A}|^2, \quad (7)$$

also equals the variance  $\sigma_{\hat{A}}^2$  of operator  $\hat{A}$ ,

$$\sigma_{\hat{A}}^2 \equiv \langle \Psi | (\hat{A} - \langle \hat{A} \rangle)^2 | \Psi \rangle \quad (8)$$

$$= \langle \Psi | (\hat{A}^2) | \Psi \rangle - |\langle \hat{A} \rangle|^2, \quad (9)$$

that is  $\sigma_A^2 = \sigma_{\hat{A}}^2$ , see Appendix. It follows that the error  $\epsilon_A(N)$  in the approximation of Eq. (4), as measured by the

standard deviation  $\sigma_A / \sqrt{N}$  of  $N$  independent samples, scales with  $N$  as

$$\epsilon_A(N) \approx \sqrt{\frac{\sigma_A^2}{N}}. \quad (10)$$

Let us analyze in which sense the above Monte Carlo sampling strategy could be of interest. The cost (i.e. computational time) of an exact contraction, Eq. (1), scales as  $O(\chi^p)$  with the bond dimension  $\chi$ . On the other hand, notice that for each specific configuration  $s$ , the contribution  $\langle \Psi | s \rangle \langle s | \hat{A} | \Psi \rangle$  to  $\langle \Psi | \hat{A} | \Psi \rangle$  consists of two tensor networks, namely one for  $\langle \Psi | s \rangle$  and another for  $\langle s | \hat{A} | \Psi \rangle$ , whose contraction can be accomplished with a cost  $O(\chi^q)$ , for some  $q < p$ , see Fig. 1. [This is also the cost of computing  $Q(s)$  and  $A(s)$  in Eq. (2)]. If the number of samples required to obtain an acceptably small error  $\epsilon_A(N)$  is  $N \approx O(\chi^{q'})$ , the use of sampling incurs a computational cost of  $O(\chi^{q+q'})$  instead of  $O(\chi^p)$ . We conclude that if  $q + q' < p$ , then (for large  $\chi$ ) the sampling strategy will have a lower computational cost than the exact contraction.

### B. Combining exact contraction with sampling: Incomplete sampling

More generally, one can consider a hybrid strategy which combines exact contraction and sampling. This is accomplished by sampling over only a subset of the  $L$  indices corresponding to the  $L$  sites of lattice  $\mathcal{L}$ , while performing an exact contraction on the remaining sites. For instance, Fig. 1(c) considers a lattice  $\mathcal{L}$  made of  $L = 6$  sites where the first three sites are being sampled, with configuration  $(s_1, s_2, s_3)$ , whereas the remaining three of sites are being addressed with an exact contraction.

If we denote by  $s^\diamond \in \mathcal{S}^\diamond$  a configuration of the  $L^\diamond$  indices to be sampled ( $L^\diamond < L$ ), then Eq. 1 is replaced with

$$\langle \Psi | \hat{A} | \Psi \rangle = \sum_{s^\diamond \in \mathcal{S}^\diamond} \langle \Psi | s^\diamond \rangle \langle s^\diamond | \hat{A} | \Psi \rangle, \quad (11)$$

We can again rewrite Eq. (11) as a probabilistic sum of an estimator  $A^\diamond(s^\diamond) \equiv \langle \Psi | s^\diamond \rangle \langle s^\diamond | \hat{A} | \Psi \rangle / |\langle \Psi | s^\diamond \rangle|^2$  according to probabilities  $Q(s^\diamond) \equiv |\langle \Psi | s^\diamond \rangle|^2$ ,

$$\langle \Psi | \hat{A} | \Psi \rangle = \sum_{s^\diamond \in \mathcal{S}^\diamond} Q(s^\diamond) A^\diamond(s^\diamond). \quad (12)$$

Similarly, we could generalize Eqs. 3-4 and apply importance sampling. We note that in this case the variance  $\sigma_{A^\diamond}^2$ , defined by

$$\sigma_{A^\diamond}^2 \equiv \sum_{s^\diamond} Q(s^\diamond) |A^\diamond(s^\diamond) - \bar{A}|^2 \quad (13)$$

$$= \sum_{s^\diamond} Q(s^\diamond) |A(s^\diamond)|^2 - |\bar{A}|^2, \quad (14)$$

might be smaller than the variance  $\sigma_{\hat{A}}^2$  of operator  $\hat{A}$  (Eq. 9), since a single incomplete sample  $s^\diamond$  corresponds to many

complete samples  $\mathbf{s}$ . [For instance, in the example of Fig. 1(c), the incomplete sample  $\mathbf{s}^\diamond = (s_1, s_2, s_3)$  corresponds to all complete samples  $\mathbf{s} = (s_1, s_2, s_3, s_4, s_5, s_6)$  that coincide with  $\mathbf{s}^\diamond$  in the first three sites.] In other words, the statistical error might be reduced. This should not come as a surprise. After all, in the extreme case where no sampling at all is performed ( $L^\diamond = 0$ ) but all indices are exactly contracted, there is no statistical error left.

### C. Markov chain Monte Carlo

In Refs. 23,24 the random configurations  $\mathbf{s}$  were generated by means of a Markov chain process based on local updates. Given a stored configuration  $\mathbf{s}$ , let us denote  $\mathbf{s}'_i$  a configuration obtained from  $\mathbf{s}$  by replacing in site  $i$  the value  $s_i$  with  $s'_i$ . Then, visiting the sites sequentially,  $i = 1, 2, \dots, L$ , in what is known as a *sweep*, a change on site  $i$  is introduced according to the Metropolis probability

$$P_{\text{change}} = \min\left[\frac{Q(\mathbf{s}'_i)}{Q(\mathbf{s})}, 1\right]. \quad (15)$$

In this way, after one sweep a new configuration  $\mathbf{s}'$  is obtained from  $\mathbf{s}$ , and by iteration a sequence of configurations

$$\mathbf{s} \rightarrow \mathbf{s}' \rightarrow \mathbf{s}'' \rightarrow \dots \quad (16)$$

is produced. However, these configurations will in general be correlated. The number  $\tau$  of sweeps required between configurations  $\mathbf{s}$  and  $\mathbf{s}'$  in order for them to be essentially independent to be independent is known as the autocorrelation time. Sweeping  $\tau$  times between samples is necessary in order for the error  $\epsilon_A(N)$  to scale as in Eq. (10), since that expression for the error assumed the samples to be independent. (If only a single sweep mediates the samples, the statistical error in Eq. (10) increases by a factor which scales as  $\tau^{1/2}$  due to autocorrelations). In addition, the first sample  $\mathbf{s}$  will be obtained after applying  $\tau'$  sweeps to some random initial configuration. The equilibration time  $\tau'$  is necessary in order to guarantee that the first sample is picked-up according to the correct probability distribution. The autocorrelation time  $\tau$  and the equilibration time  $\tau'$  are known to diverge with systems size  $L$  for critical systems.

Large equilibration and autocorrelation times, e.g. near or at a critical point, increase the cost of simulations. This increase can be prevented if somehow independent configurations  $\mathbf{s}$  can be directly generated according to probabilities  $Q(\mathbf{s})$ . In section IV we show how this is possible for a specific class of tensor networks, namely unitary tensor networks, which are introduced next.

### III. SAMPLING OF UNITARY TENSOR NETWORKS

Let us specialize to the particular case of unitary tensor networks, namely tensor networks that are based on a unitary quantum circuit. Examples include the MERA and unitary

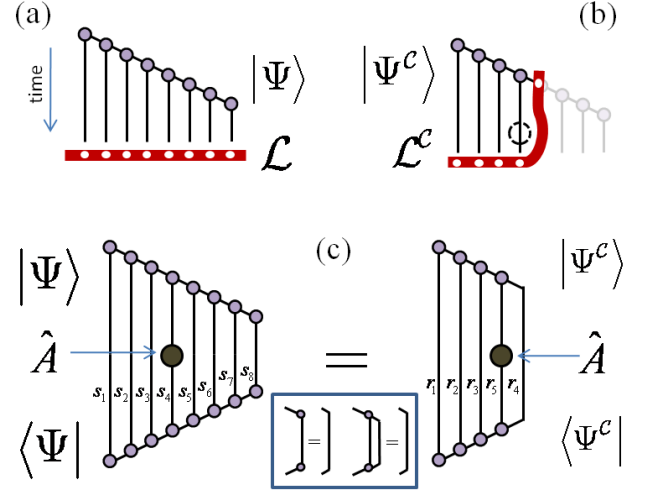


FIG. 2: (Color online) Sampling in a unitary matrix product state (uMPS). (a) uMPS for a state  $|\Psi\rangle$  of lattice  $\mathcal{L}$ . Notice the (fictitious) time direction, which provides each tensor with a sense of which indices are incoming and which are outgoing. (b) The past causal cone  $\mathcal{C}$  of a local operator  $\hat{A}$  acting on a single site of  $\mathcal{L}$  (denoted by a discontinuous circle) defines an effective lattice  $\mathcal{L}^C$ , which is found in state  $|\Psi^C\rangle$ . Notice that the effective lattice  $\mathcal{L}^C$  is made of two types of sites, namely sites already present in the original lattice  $\mathcal{L}$  and one site not present in  $\mathcal{L}$ , with  $d$ -dimensional and  $\chi$ -dimensional vector spaces, respectively. (c) Tensor networks representing  $\langle \Psi | \hat{A} | \Psi \rangle$  and  $\langle \Psi^C | \hat{A} | \Psi^C \rangle$ . The inset shows unitarity reductions [Eq. (17)] used to transform  $\langle \Psi | \hat{A} | \Psi \rangle$  into  $\langle \Psi^C | \hat{A} | \Psi^C \rangle$ .

versions of MPS (with open boundary conditions) and TTN, which we will refer as uMPS and uTTN<sup>32</sup>.

Unitary tensor networks are special in that each tensor  $u$  is constrained to be unitary/isometric. Figs. 2 and 3 exemplify the discussion for uMPS and uTTN respectively. Specifically, we first note that in one such tensor network there is a well-defined direction of time throughout, see e.g. Figs. 2(a) and 3(a). Each index of a tensor  $u$  is either an incoming index (if time flows towards the tensor) or an outgoing index (if time flows away from the tensor). The constraint on  $u$  can be expressed in the following way. Let us group all incoming indices of  $u$  into a composite incoming index  $\alpha$  and all outgoing indices of  $u$  into a composite outgoing index  $\beta$ , so that tensor  $u$  becomes a matrix  $u_{\beta\alpha}$ . Then the unitary/isometric constraint on  $u$  reads

$$\sum_{\beta} (u^\dagger)_{\alpha\beta} u_{\beta\alpha'} = \delta_{\alpha\alpha'}. \quad (17)$$

A direct implication of this property is that the tensor network corresponding to the expectation value  $\langle \Psi | \hat{A} | \Psi \rangle$  can be replaced with a simplified tensor network where the pairs of tensors  $(u, u^\dagger)$  outside the so-called past causal cone  $\mathcal{C}$  of  $\hat{A}$  have been removed, see Figs. 2(c) and 3(c). This new tensor network can be interpreted to represent the expectation value  $\langle \Psi^C | \hat{A} | \Psi^C \rangle$  of the local operator  $\hat{A}$  on a state  $|\Psi^C\rangle \in \mathbb{V}_{\mathcal{L}^C}$  of an effective lattice  $\mathcal{L}^C$  defined by the causal cone  $\mathcal{C}$  of the operator  $\hat{A}$ , see Figs. 2(b) and 3(b), where by construction



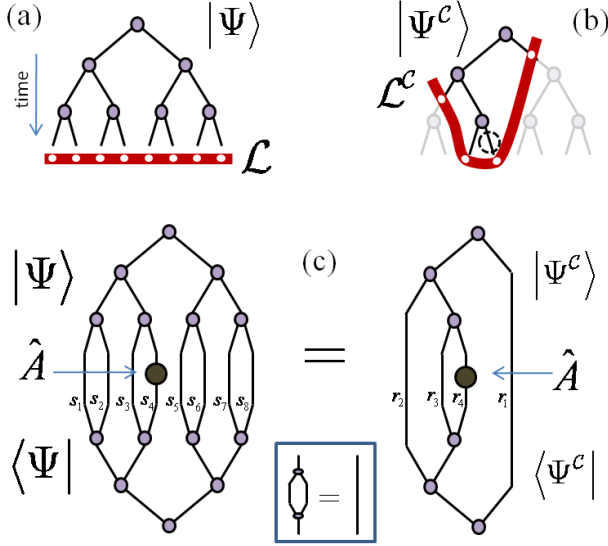


FIG. 3: (Color online) Sampling in a unitary tree tensor network (uTTN). (a) uTTN for a state  $|\Psi\rangle$  of lattice  $\mathcal{L}$ . (b) Effective lattice  $\mathcal{L}^c$ . (c) Tensor networks for  $\langle\Psi|O|\Psi\rangle$  and  $\langle\Psi^c|A|\Psi^c\rangle$ . The inset shows a reduction due to the unitary constrain of tensors in the uTTN.

$\langle\Psi|\hat{A}|\Psi\rangle = \langle\Psi^c|\hat{A}|\Psi^c\rangle$ . The effective lattice  $\mathcal{L}^c$  is made of  $L^c$  sites that can be of two types: those already contained in the original lattice  $\mathcal{L}$ , which are described by a  $d$ -dimensional vector space, and those which did not belong to  $\mathcal{L}$ , which are described by a  $\chi$ -dimensional vector space. We use  $\mathbf{r} = (r_1, r_2, \dots, r_{L^c})$  to denote a configuration of the effective lattice  $\mathcal{L}^c$ , and  $|\mathbf{r}\rangle \equiv |r_1\rangle \otimes |r_2\rangle \otimes \dots \otimes |r_{L^c}\rangle$  the corresponding product vector, where for some sites  $r_i = 1, 2, \dots, d$  and for some others  $r_i = 1, 2, \dots, \chi$ . We denote  $\mathcal{R}$  the set of all configurations  $\mathbf{r}$ .

The exact contraction of the tensor network corresponding to  $\langle\Psi^c|\hat{A}|\Psi^c\rangle$ , may still be very expensive and again we might be interested in exploring the use of sampling to lower the computational cost. For that purpose, we repeat the discussion in section II. First we write the expectation value  $\langle\Psi|\hat{A}|\Psi\rangle$  as

$$\langle\Psi|\hat{A}|\Psi\rangle = \sum_{\mathbf{r} \in \mathcal{R}} \langle\Psi^c|\mathbf{r}\rangle \langle\mathbf{r}|\hat{A}|\Psi^c\rangle, \quad (18)$$

see Fig. 4 for uMPS and uTTN. Then we rewrite Eq. (18) in terms of the estimator  $A^c(\mathbf{r}) \equiv \langle\mathbf{r}|\hat{A}|\Psi^c\rangle / \langle\mathbf{r}|\Psi^c\rangle$  and probabilities  $P(\mathbf{r}) \equiv |\langle\mathbf{r}|\Psi^c\rangle|^2$ ,

$$\langle\Psi|\hat{A}|\Psi\rangle = \sum_{\mathbf{r} \in \mathcal{R}} P(\mathbf{r}) A^c(\mathbf{r}). \quad (19)$$

We can again limit the sum over configurations  $\mathbf{r}$  to a subset  $\tilde{\mathcal{R}}$  containing just  $N$  configurations which, when chosen from  $\mathcal{R}$  randomly according to the probabilities  $P(\mathbf{r})$ , results in

$$\langle\Psi|\hat{A}|\Psi\rangle \approx \frac{1}{N} \sum_{\mathbf{r} \in \tilde{\mathcal{R}}} A^c(\mathbf{r}). \quad (20)$$

The error in the approximation scales with  $N$  as in Eq. (10).

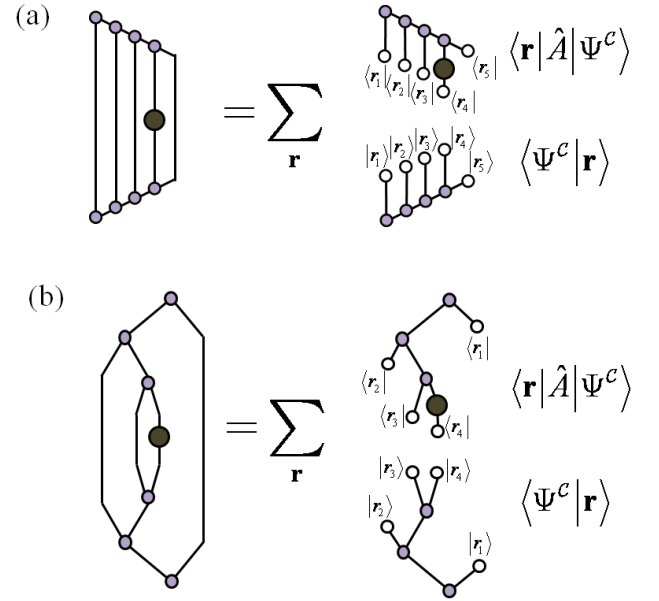


FIG. 4: (Color online) Graphical representation of  $\langle\Psi^c|\hat{A}|\Psi^c\rangle = \sum_{\mathbf{r} \in \mathcal{R}} \langle\Psi^c|\mathbf{r}\rangle \langle\mathbf{r}|\hat{A}|\Psi^c\rangle$ . In (a), the original state  $|\Psi\rangle$  was represented with an uMPS, see Fig. 2. In (b), the original state  $|\Psi\rangle$  was represented with an uTTN, see Fig. 3. However, in both cases the state  $|\Psi^c\rangle$  is represented by an uMPS that runs through the causal cone.

#### IV. PERFECT SAMPLING

In this section we describe how to randomly draw configurations  $\mathbf{r}$  according to probability  $P(\mathbf{r})$  in a unitary tensor network. We refer to this scheme as perfect sampling because, in contrast with Markov chain Monte Carlo, the present scheme produces perfectly uncorrelated samples. We will also refer to this scheme as complete perfect sampling, to distinguish it from the incomplete perfect sampling scheme discussed in the next section, where sampling is performed only on a subset of sites.

##### A. Algorithm

Recall that as a quantum circuit, the tensor network is equipped with a notion of (fictitious) time. From now on we assume that the labeling of the sites in the effective lattice  $\mathcal{L}^c$  has been chosen so as to progress forward with respect to this notion of time. Thus, site 1 corresponds to the earliest time, site 2 corresponds to a later time, and so on, until site  $L^c$  corresponds to the latest time (when two sites correspond to the same time, e.g. sites 4 and 5 in Fig. 4 (a), we order them arbitrarily).

Our *perfect sampling* algorithm consists of sequentially computing a series of conditional single-site density matrices  $\{\rho_1, \rho_2(r_1), \dots\}$  and conditional single-site probabilities  $\{P(r_1), P(r_2|r_1), \dots\}$ . First we compute the reduced density

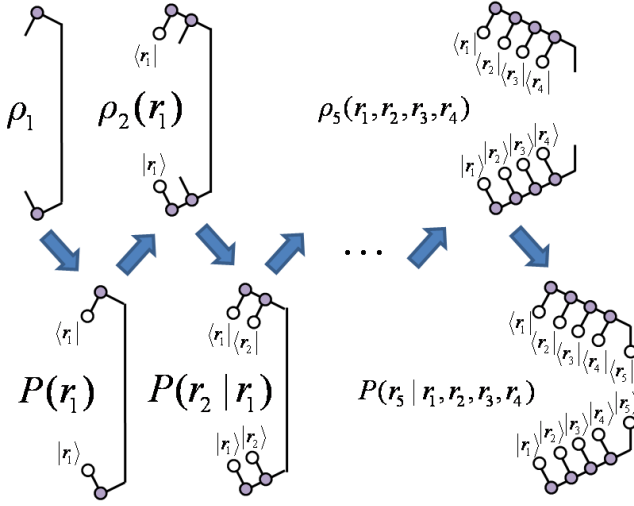


FIG. 5: (Color online) Perfect sampling with a uMPS. The figure shows a sequence of the tensor networks corresponding (up to a proportionality constant) to  $\rho_1$ ,  $P(r_1)$ ,  $\rho_2(r_1)$ ,  $P(r_2|r_1)$ , and so on, see Eqs. (21–28). Importantly, all these tensor networks can be contracted with a cost that scales as  $O(\chi^2)$  with the bond dimension  $\chi$ , and are therefore computational less expensive than an exact contraction, which has cost  $O(\chi^3)$ .

matrix  $\rho_1$  for site 1 exactly, i.e. without sampling,

$$\rho_1 \equiv \text{tr}_{2\dots L^c} \{ |\Psi^c\rangle \langle \Psi^c| \} \quad (21)$$

from which we can compute the probabilities

$$P(r_1) \equiv \langle r_1 | \rho_1 | r_1 \rangle. \quad (22)$$

We can then randomly choose a value for  $r_1$  according to probability  $P(r_1)$ , and compute (exactly) the conditional reduced density matrix  $\rho_2(r_1)$  for site 2, which is obtained from the state  $\langle r_1 | \Psi^c \rangle$  of sites 2 to  $L^c$ ,

$$\rho_2(r_1) \equiv \frac{1}{P(r_1)} \text{tr}_{3\dots L^c} \{ \langle r_1 | \Psi^c \rangle \langle \Psi^c | r_1 \rangle \}. \quad (23)$$

Again, we can use the reduced density matrix to compute the conditional probabilities

$$P(r_2|r_1) \equiv \langle r_2 | \rho_2(r_1) | r_2 \rangle, \quad (24)$$

and we can therefore randomly select a value of  $r_2$  according to probabilities  $P(r_2|r_1)$ . Let us notice at this point that so far we have randomly chosen values for  $r_1$  and  $r_2$  according to the probability

$$P(r_1, r_2) = P(r_1)P(r_2|r_1) = ||\langle r_1, r_2 | \Psi^c \rangle||^2. \quad (25)$$

We can now iterate the above process, that is, compute the conditional density matrix

$$\rho_3(r_1, r_2) \equiv \frac{1}{P(r_1, r_2)} \text{tr}_{4\dots L^c} \{ \langle r_1, r_2 | \Psi^c \rangle \langle \Psi^c | r_1, r_2 \rangle \} \quad (26)$$

and the conditional probabilities

$$P(r_3|r_1, r_2) \equiv \langle r_3 | \rho_3(r_1, r_2) | r_3 \rangle, \quad (27)$$

and so on for the rest of sites in the effective lattice  $\mathcal{L}^c$ . In this way, and since

$$P(\mathbf{r}) = P(r_1)P(r_2|r_1) \cdots P(r_{L^c}|r_1, r_2, \dots, r_{L^c-1}), \quad (28)$$

we end up indeed randomly choosing a configuration  $\mathbf{r} = (r_1, r_2, \dots, r_{L^c})$  with probability given precisely by  $P(\mathbf{r}) \equiv |\langle \mathbf{r} | \Psi^c \rangle|^2$ .

Fig. 5 illustrates the sequence of computations in the case of a one-site operator  $\hat{A}$  specifically for a uMPS, assuming as in Figs. 2 and 4(a) that the operator  $\hat{A}$  is supported on the fourth site of the original chain. This algorithm is similar to one used for thermal state sampling with MPS<sup>25</sup> described in Ref. 28. Analogous computations for a uTTN are very similar, since the causal cone of a single-site operator  $\hat{A}$  is described also by a uMPS, see Fig. 3(b). For the case of a MERA, more details on the implementation of Eqs. (21–28) can be found in Ref. 29.

A key point is that, for unitary tensor networks such as uMPS, uTTN, and MERA, the computational cost of generating the above sequence of density matrices and probabilities often does not exceed (to leading order in  $\chi$  and effective size  $L^c$ ) the cost of a single sweep in Markov chain Monte Carlo<sup>33</sup>.

## B. Benchmark

To illustrate the performance of the perfect sampling scheme and compare it to Markov chain Monte Carlo, we have considered a duly optimized uMPS for the ground state  $|\Psi\rangle$  of the quantum Ising model with critical transverse magnetic field,

$$\hat{H}_{\text{Ising}} \equiv - \sum_{\langle i,j \rangle} \hat{\sigma}_i^z \hat{\sigma}_j^z - \sum_i \hat{\sigma}_i^x, \quad (29)$$

on an open chain of  $L$  spins.<sup>35</sup> The two sampling schemes are then used in order to compute the expectation value of local operators.

Fig. 6(a) and (b) show a history of 150 configurations of a chain of  $L = 50$  spins obtained with perfect sampling and Markov chain Monte Carlo, respectively. The existence of correlations in the second case is manifest.

Fig. 6(c) and (d) show the error in the expectation value  $\langle \Psi | \hat{\sigma}_{25}^z | \Psi \rangle$  and  $\langle \Psi | \hat{\sigma}_{25}^x | \Psi \rangle$  for the local operators  $\sigma^z$  and  $\sigma^x$  on site 25, as a function of the number of samples  $N$ . In both cases, the effect of autocorrelations in Markov chain Monte Carlo results in an error larger than the error obtained with perfect sampling, which is given by Eq. (10). The ratio between statistical errors, as given in terms of the autocorrelation time  $\tau$  by  $\sqrt{2\tau + 1}$ , is seen to depend on the choice of local operator – this autocorrelation time is larger for  $\langle \Psi | \hat{\sigma}_{25}^z | \Psi \rangle$  than for  $\langle \Psi | \hat{\sigma}_{25}^x | \Psi \rangle$ .

Finally, Fig. 6 (e) and (f) explore the autocorrelation time  $\tau$  for  $\hat{\sigma}^z$  as a function of the size  $L$  of the spin chain. In particular, Fig. 6 (f) reveals that  $\tau$  grows linearly in  $L$ . This means<sup>36</sup>

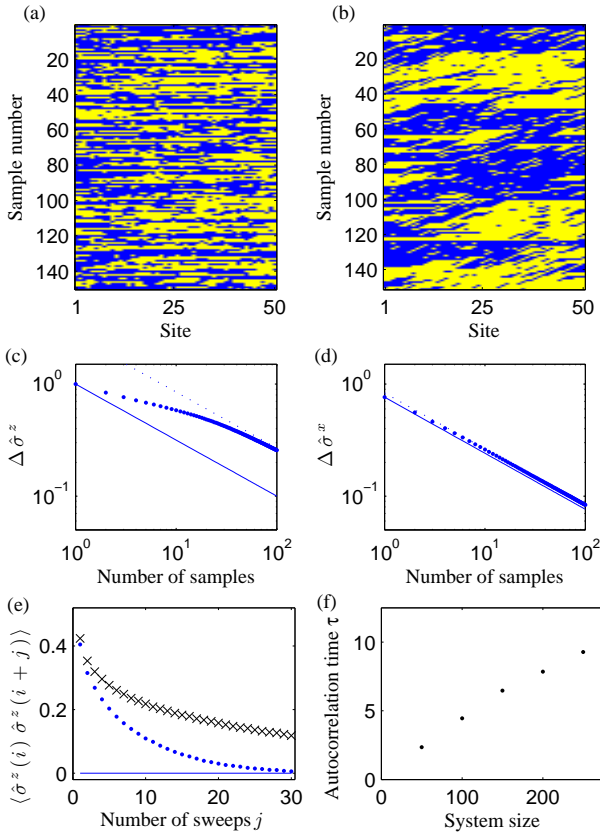


FIG. 6: (Color online) Sampling of the ground state of the critical transverse Ising model in the  $z$  basis. Comparison between configurations obtained using (a) the presented perfect sampling scheme and (b) a Markov chain scheme (single sweep) on 50 sites. Blue sites represent spin up and yellow for spin down. The correlations between configurations obtained using a Markov chain scheme are evidenced by the appearance of domains of well defined color that extend vertically. In (c) we have calculated the expected statistical error on the estimate of  $\langle \hat{\sigma}^z \rangle$  for the perfect sampling (blue line) and Markov chain sampling (blue dots). While with perfect sampling the error decreases with the usual  $N^{-1/2}$  factor, correlations between subsequent samples increase the error on the estimate in the Markov chain scheme. In (d) we plot the same for  $\langle \hat{\sigma}^x \rangle$  by projecting all the spins into the  $x$  basis. In this case the Markov scheme used utilizes a 2-site update so as to be compatible with the wave-function symmetry<sup>34</sup>. In (e) we present the correlations on the centre site (in the  $z$  basis) after  $j$  Markov chain sweeps using  $10^6$  samples for 50 sites (blue dots) and 250 sites (black crosses). In the perfect sampling scheme (blue line), there are no correlations between configurations. In (f) we plot the estimated autocorrelation time for different system sizes.

that in order to achieve a fixed accuracy in  $\langle \Psi | \hat{\sigma}_{L/2}^z | \Psi \rangle$ , the number of samples  $N$  with Markov chain Monte Carlo has to grow linearly in  $L$ , whereas a constant number of samples is enough with perfect sampling.

It is important to stress, however, that the Markov chain Monte Carlo update scheme discussed here, based on single spin updates, is used as a reference only – more sophisticated Markov chain Monte Carlo schemes, based e.g. on global spin updates, could lead to smaller autocorrelation times.

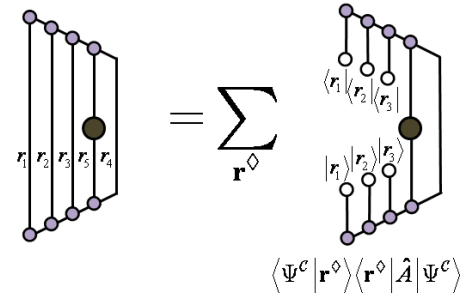


FIG. 7: (Color online) Graphical representation of  $\langle \Psi^C | \hat{A} | \Psi^C \rangle = \sum_{\mathbf{r}^\diamond \in \mathcal{R}^\diamond} \langle \Psi^C | \mathbf{r}^\diamond \rangle \langle \mathbf{r}^\diamond | \hat{A} | \Psi^C \rangle$  for a uMPS, to be compared with Fig. 4(a). Notice that sampling does not affect two of the indices, over which an exact contraction is still performed.

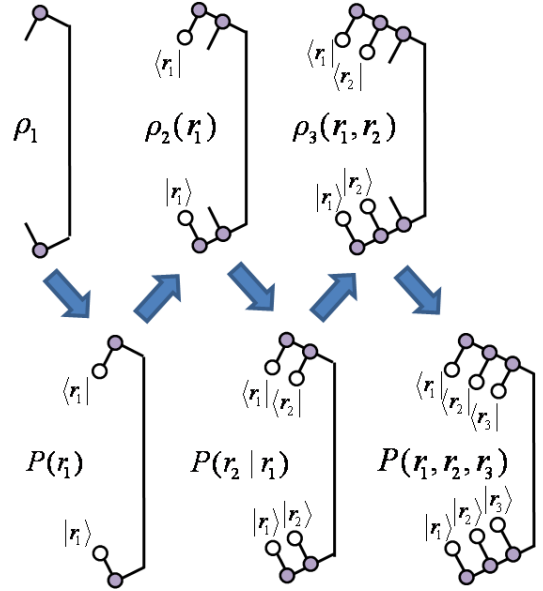


FIG. 8: (Color online) Incomplete perfect sampling with a uMPS. The figure shows a complete sequence of the tensor networks corresponding (up to a proportionality constant) to  $\rho_1$ ,  $P(r_1)$ ,  $\rho_2(r_2)$ ,  $P(r_2|r_1)$ ,  $\rho_3(r_1, r_2)$  and  $P(r_1, r_2, r_3)$  necessary in order to generate a configuration  $\mathbf{r}^\diamond = (r_1, r_2, r_3)$  with probability  $P(\mathbf{r}^\diamond) = |\langle \Psi^C | \mathbf{r}^\diamond \rangle|^2$ . Notice that the cost still scales as  $O(\chi^2)$ , as in the complete (perfect) sampling scheme.

## V. INCOMPLETE PERFECT SAMPLING

So far we have considered perfect sampling over the whole causal cone, that is, over the indices associated to all the sites of the effective lattice  $\mathcal{L}^C$ . However, it is also possible to use an incomplete perfect sampling scheme, which combines perfect sampling over most of the sites of  $\mathcal{L}^C$  and an exact contraction over a small set of sites, without altering the scaling  $O(\chi^q)$  of the cost of a single sample. Because we are sampling over fewer indices, we can expect a decrease in the statistical error with little change in the cost. In some cases the reduction in statistical uncertainty can be dramatic.

### A. Incomplete perfect sampling scheme

The incomplete perfect sampling scheme is illustrated in Fig. 7 for a uMPS. The first step is to rewrite the expectation value  $\langle \Psi^C | \hat{A} | \Psi^C \rangle = \langle \Psi | \hat{A} | \Psi \rangle$  as

$$\langle \Psi | \hat{A} | \Psi \rangle = \sum_{\mathbf{r}^\diamond \in \mathcal{R}^\diamond} \langle \Psi^C | \mathbf{r}^\diamond \rangle \langle \mathbf{r}^\diamond | \hat{A} | \Psi^C \rangle, \quad (30)$$

where  $\mathcal{R}^\diamond$  is the set of incomplete configurations  $\mathbf{r}^\diamond \equiv (r_1, r_2, \dots, r_{L^\diamond})$ , where  $L^\diamond$  is the number of sites over which sampling takes place, with  $L^\diamond < L^C$ . For the case of the uMPS illustrated in Fig. 7, one can perform an exact contraction on two sites of  $\mathcal{L}^C$ , namely the site on which the local operator  $\hat{A}$  is supported and the effective,  $\chi$ -dimensional site corresponding to the bond index of the uMPS. Notice that now the term  $\langle \Psi^C | \mathbf{r}^\diamond \rangle \langle \mathbf{r}^\diamond | \hat{A} | \Psi^C \rangle$  does not factorize into two terms, since  $\langle \mathbf{r}^\diamond | \Psi^C \rangle$  and  $\langle \mathbf{r}^\diamond | \hat{A} | \Psi^C \rangle$  are no longer complex numbers but  $d\chi$ -dimensional vectors.

We can still rewrite Eq. (30) as a probabilistic sum of an estimator  $A^\diamond(\mathbf{r}^\diamond) \equiv \langle \Psi^C | \mathbf{r}^\diamond \rangle \langle \mathbf{r}^\diamond | \hat{A} | \Psi^C \rangle / |\langle \Psi^C | \mathbf{r}^\diamond \rangle|^2$  according to probabilities  $P(\mathbf{r}^\diamond) \equiv |\langle \Psi^C | \mathbf{r}^\diamond \rangle|^2$ ,

$$\langle \Psi | \hat{A} | \Psi \rangle = \sum_{\mathbf{r}^\diamond \in \mathcal{R}^\diamond} P(\mathbf{r}^\diamond) A^\diamond(\mathbf{r}^\diamond), \quad (31)$$

limiting the sum over configurations  $\mathbf{r}^\diamond$  to a subset  $\tilde{\mathcal{R}}^\diamond$  containing just  $N$  configurations, and use (perfect) importance sampling to obtain the estimate

$$\langle \Psi | \hat{A} | \Psi \rangle \approx \frac{1}{N} \sum_{\mathbf{r}^\diamond \in \tilde{\mathcal{R}}^\diamond} A^\diamond(\mathbf{r}^\diamond). \quad (32)$$

An important difference between the incomplete perfect sampling scheme and the complete perfect sampling scheme of Eqs. (18–20) is that the estimator  $A^\diamond$ , whose mean is  $\bar{A}^\diamond = \langle \Psi | \hat{A} | \Psi \rangle$  as indicated in Eq. (31), has a variance  $\sigma_{A^\diamond}^2$ ,

$$\sigma_{A^\diamond}^2 \equiv \sum_{\mathbf{r}^\diamond \in \mathcal{R}^\diamond} P(\mathbf{r}^\diamond) |A^\diamond(\mathbf{r}^\diamond) - \bar{A}^\diamond|^2 \quad (33)$$

$$= \sum_{\mathbf{r}^\diamond \in \mathcal{R}^\diamond} P(\mathbf{r}^\diamond) |A^\diamond(\mathbf{r}^\diamond)|^2 - |\bar{A}^\diamond|^2, \quad (34)$$

that is no longer necessarily equal to the variance  $\sigma_{\hat{A}}^2$  of Eq. (9), but is instead upper bounded by it,  $\sigma_{A^\diamond}^2 \leq \sigma_{\hat{A}}^2$ , see the Appendix. In other words, the error  $\epsilon_{A^\diamond}(N)$  in the approximation of Eq. (32), given by

$$\epsilon_{A^\diamond}(N) \approx \sqrt{\frac{\sigma_{A^\diamond}^2}{N}}, \quad (35)$$

can be smaller than the error  $\epsilon_A(N)$  of a complete sampling scheme.

### B. Algorithm

We have implemented the incomplete perfect sampling scheme in conjunction with the complete perfect sampling

scheme described in section IV. We notice, however, that incomplete sampling can also be incorporated into Markov chain Monte Carlo.

As in section IV, we proceed by constructing a sequence of conditional single-site reduced density matrices  $\{\rho_1, \rho_2(r_1), \dots\}$  and conditional probabilities  $\{P(r_1), P(r_2|r_1), \dots\}$ . However, in this occasion the sequence concludes at site  $L^\diamond$ , after which we can already evaluate the estimator  $A^\diamond(\mathbf{r}^\diamond)$ . This is illustrated for the case of a uMPS in Fig. 8, which is to be compared with Fig. 5.

### C. Benchmark

As in section IV, we use sampling to compute the expectation value of local observables from a uMPS with  $\chi = 30$  that has been previously optimized to approximate the ground state of the quantum Ising chain at criticality, Eq. (29). The exact structure that we sample can be seen in Fig. 7. Figure 9 shows the sampling error, as a function of the number of samples  $N$ , in the computation of  $\langle \Psi | \hat{\sigma}_{25}^z | \Psi \rangle$  and  $\langle \Psi | \hat{\sigma}_{25}^x | \Psi \rangle$  in a chain of  $L = 50$  spins. The error is seen to depend on two factors. On the one hand, it depends on which operator ( $\hat{\sigma}^z$  or  $\hat{\sigma}^x$ ) is being measured, as it did in section IV. In addition, now it also drastically depends on which product basis  $\{|\mathbf{r}^\diamond\rangle\}$  is used. In particular, we see that a very substantial reduction of sampling error, of seven orders of magnitude, is obtained by measuring on the  $x$  basis while computing  $\langle \Psi | \hat{\sigma}_{25}^z | \Psi \rangle$ . It should be noted that the two-site Markov chain update scheme used for the  $x$ -basis calculations,<sup>34</sup> although appears competitive, is more computationally demanding than the perfect sampling scheme and runs approximately 2–3 times slower.

## VI. COMPUTATIONAL COSTS

For completeness, we include a brief summary of the computational costs incurred in extracting, from a given unitary tensor network, the expectation value of a local operator by using (i) exact contraction, (ii) Markov chain Monte Carlo and (iii) a perfect sampling scheme. For simplicity, we consider only one-site local operators. The scaling of the costs in the bond dimension  $\chi$  is presented in Table I. We emphasize that in the sampling schemes, we only consider the cost of obtaining one sample. A fair comparison of costs with an exact contraction should also take into account the number of samples required in order to approximate the exact result with some pre-agreed accuracy.

The table shows that for both a uMPS and the MERA, the cost of Markov chain Monte Carlo and perfect sampling scale with the same power. Instead, for the uTTN, the cost of Markov chain Monte Carlo is one power smaller than that of perfect sampling. [The same would happen with uMPS if the local dimension of each site was also  $\chi$ ]. More significant speed-ups can be seen with the MERA, both for the computation of two-point correlators, and in systems in two dimensions (not in the Table), where sampling techniques to increase computational



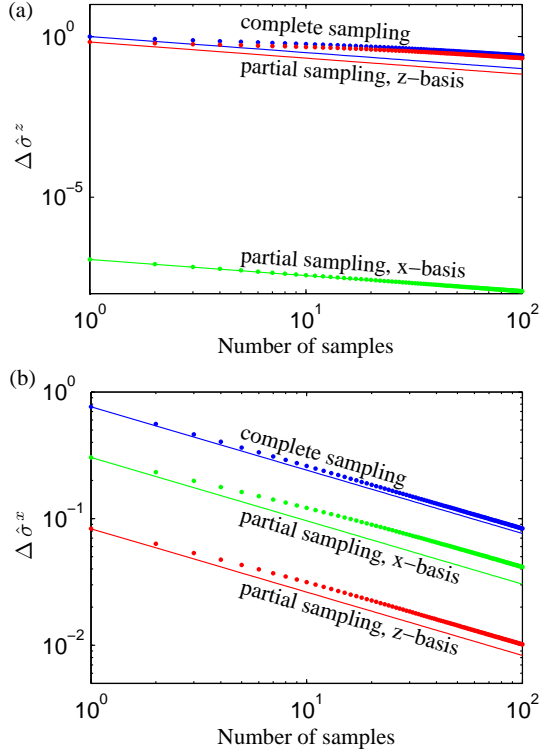


FIG. 9: (Color online) Sampling errors with the incomplete perfect sampling scheme for a 50 site critical Ising chain, using both perfect sampling (continuous lines) and Markov chain Monte Carlo sampling (dots). (a) Sampling errors in the computation of  $\langle \Psi | \hat{\sigma}_{25}^z | \Psi \rangle$ . With perfect sampling, errors in the incomplete perfect sampling scheme are upper-bounded by the errors in a complete sampling scheme, as proven in the Appendix. Interestingly, for estimates of  $\langle \hat{\sigma}_z \rangle$  the incomplete perfect sampling scheme obtains an error  $10^{-7}$  times smaller by measuring in the  $x$  basis on sites  $1, 2, \dots, L^\circ$ . (b) Sampling errors in the computation of  $\langle \Psi | \hat{\sigma}_{25}^x | \Psi \rangle$ . Again, the errors with incomplete perfect sampling are smaller than those with complete perfect sampling, and depend on the choice of product basis.

efficiency are required most. The authors present an in-depth analysis of perfect sampling with the MERA in Ref. 29.

A further remark is in order. The above analysis assumes that a tensor network has been provided in a unitary circuit form. In particular, the costs in Table I do not include operations such as converting a non-unitary version of the tensor network into its unitary form (typically through the QR-decomposition). In particular, the cost of QR-decompositions required to turn an MPS into a uMPS scales as  $\mathcal{O}(\chi^3)$  – that is, the same scaling as an exact contraction. What is then the practical interest in a perfect sampling scheme for a uMPS? On the one hand, the uMPS might conceivably have been generated through some procedure (e.g. along the lines of the algebraic Bethe Ansatz MPS constructions described in Ref. 30), with a cost  $\mathcal{O}(\chi^2)$  (notice that a uMPS tensor only contains  $\mathcal{O}(\chi^2)$  coefficients). In this case, the perfect sampling scheme would allow for a very efficient, approximate evaluation of expectation values without increasing this cost. On

TABLE I: The leading-order costs of contracting unitary tensor networks with and without sampling techniques, with the goal of estimating the expectation value of a one-site operator. For the MERA we have also included the cost calculating arbitrary (long-range) two-point correlators.<sup>29</sup>

Tensor network	Exact contraction	Markov-chain MC	Perfect sampling
uMPS (open BC)	$\mathcal{O}(\chi^3)$	$\mathcal{O}(\chi^2)$	$\mathcal{O}(\chi^2)$
uTTN (binary)	$\mathcal{O}(\chi^4)$	$\mathcal{O}(\chi^2)$	$\mathcal{O}(\chi^3)$
MERA (1D binary)	$\mathcal{O}(\chi^9)$	$\mathcal{O}(\chi^5)$	$\mathcal{O}(\chi^5)$
→ 2-point correlators	$\mathcal{O}(\chi^{12})$	$\mathcal{O}(\chi^7)$	$\mathcal{O}(\chi^8)$

the other hand, although we have focused our analysis on the evaluation of local expectation values, more complex tasks involving a uMPS, such as the computation of entanglement entropy, can exploit the perfect sampling schemes presented in this paper at a cost significantly lower than that of an exact contraction (see e.g. Ref. 31).

## VII. CONCLUSIONS

We have explained how to perform Monte Carlo sampling on unitary tensor networks such as the MERA, uMPS and uTTN. In order to compute the expectation value  $\langle \Psi | \hat{A} | \Psi \rangle$  of a local operator  $\hat{A}$ , sampling is performed on the past causal cone  $\mathcal{C}$  of operator  $\hat{A}$ . In addition, by exploiting the unitary character of the tensors, it is possible to directly sample configurations  $\mathbf{r}$  of the causal cone according to their weight in the wave-function, resulting in uncorrelated samples and thus avoiding the equilibration and autocorrelation times of Markov chain Monte Carlo schemes. This last property makes the perfect sampling scheme particularly interesting to study critical systems.

In principle, one can also proceed as in Eqs. (21–28) for non-unitary tensor networks, e.g. PEPS, and obtain perfect sampling. However, in non-unitary tensor networks the cost of computing e.g.  $\rho_1$  is already the same as that of computing the expectation value  $\langle \Psi | \hat{A} | \Psi \rangle$  without sampling. Therefore perfect sampling in non-unitary tensor networks seems to be of very limited interest.

Here we have only considered sampling in the context of computing expectation values. However, the same approach can also be applied in order to optimize the variational ansatz, as discussed in full detail in Ref. 29 for the MERA.

The authors thank Glen Evenbly for useful discussions. Support from the Australian Research Council (FF0668731, DP0878830, DP1092513), the visitor programme at Perimeter Institute, NSERC and FQRNT is acknowledged.

## Appendix A: Variance with complete and incomplete sampling

Given a vector  $|\Psi\rangle \in \mathbb{V}_{\mathcal{L}}$  and a local operator  $\hat{A}$ , the expectation value of  $\hat{A}$  is given by  $\langle\Psi|\hat{A}|\Psi\rangle$  and its variance is

$$\sigma_{\hat{A}}^2 \equiv \langle\Psi| \left( |\hat{A} - \langle\Psi|\hat{A}|\Psi\rangle|^2 \right) |\Psi\rangle \quad (\text{A1})$$

$$= \langle\Psi| \left( |\hat{A}|^2 \right) |\Psi\rangle - |\langle\Psi|\hat{A}|\Psi\rangle|^2. \quad (\text{A2})$$

### 1. Mean and variance with complete sampling

Consider the complex random variable  $(A(s), P(s))$ , where  $A(s)$  is the estimator

$$A(s) \equiv \frac{\langle\Psi|s\rangle\langle s|\hat{A}|\Psi\rangle}{\langle\Psi|s\rangle\langle s|\Psi\rangle} = \frac{\langle s|\hat{A}|\Psi\rangle}{\langle s|\Psi\rangle} \quad (\text{A3})$$

and  $Q(s)$  is the probability

$$Q(s) \equiv \langle\Psi|s\rangle\langle s|\Psi\rangle. \quad (\text{A4})$$

Here  $\{|s\rangle\}$  denotes an orthonormal basis in the vector space  $\mathbb{V}_{\mathcal{L}}$ . Notice that  $\sum_s |s\rangle\langle s|$  is a resolution of the identity in  $\mathbb{V}_{\mathcal{L}}$  and therefore  $\sum_s Q(s) = \langle\Psi|\Psi\rangle = 1$ .

The mean  $\bar{A}$  is given by the expectation value  $\langle\Psi|\hat{A}|\Psi\rangle$ ,

$$\begin{aligned} \bar{A} &\equiv \sum_s Q(s)A(s) = \sum_s \langle\Psi|s\rangle\langle s|\Psi\rangle \frac{\langle\Psi|s\rangle\langle s|\hat{A}|\Psi\rangle}{\langle\Psi|s\rangle\langle s|\Psi\rangle} \\ &= \sum_s \langle\Psi|s\rangle\langle s|\hat{A}|\Psi\rangle = \langle\Psi|\hat{A}|\Psi\rangle. \end{aligned} \quad (\text{A5})$$

In turn, its variance  $\sigma_{\hat{A}}^2$ ,

$$\sigma_{\hat{A}}^2 \equiv \sum_s Q(s)|A(s) - \bar{A}|^2 \quad (\text{A6})$$

$$= \sum_s Q(s)|A(s)|^2 - |\bar{A}|^2, \quad (\text{A7})$$

equals the variance  $\sigma_{\hat{A}}^2$  of operator  $\hat{A}$ , as can be seen from

$$\begin{aligned} \sum_s Q(s)|A(s)|^2 &= \sum_s \langle\Psi|s\rangle\langle s|\Psi\rangle \frac{\langle\Psi|\hat{A}^\dagger|s\rangle\langle s|\hat{A}|\Psi\rangle}{\langle\Psi|s\rangle\langle s|\Psi\rangle} \\ &= \sum_s \langle\Psi|\hat{A}^\dagger|s\rangle\langle s|\hat{A}|\Psi\rangle = \langle\Psi| \left( |\hat{A}|^2 \right) |\Psi\rangle. \end{aligned} \quad (\text{A8})$$

### 2. Mean and variance with incomplete sampling

Consider now a new complex random variable  $(A(s), Q(s))$ , where  $A(s)$  is the estimator

$$A(s) \equiv \frac{\langle\Psi|\pi(s)\hat{A}|\Psi\rangle}{\langle\Psi|\pi(s)|\Psi\rangle} \quad (\text{A9})$$

and  $Q(s)$  is the probability

$$Q(s) \equiv \langle\Psi|\pi(s)|\Psi\rangle. \quad (\text{A10})$$

Here  $\{\pi(s)\}$  denotes a complete set of projectors on the vector space  $\mathbb{V}_{\mathcal{L}}$ , that is  $\pi(s)^2 = \pi(s)$ , and  $\sum_s \pi(s)$  is a resolution of the identity in  $\mathbb{V}_{\mathcal{L}}$ , so that  $\sum_s Q(s) = \langle\Psi|\Psi\rangle = 1$ . Notice that if all the projectors  $\pi(s)$  have rank one, then we recover the situation analyzed in the previous subsection. Notice also that this more general setting includes the case addressed in Sect. V in the context of incomplete sampling.

The mean  $\bar{A}$  is again given by the expectation value  $\langle\Psi|\hat{A}|\Psi\rangle$ ,

$$\begin{aligned} \bar{A} &\equiv \sum_s Q(s)A(s) = \sum_s \langle\Psi|\pi(s)|\Psi\rangle \frac{\langle\Psi|\pi(s)\hat{A}|\Psi\rangle}{\langle\Psi|\pi(s)|\Psi\rangle} \\ &= \sum_s \langle\Psi|\pi(s)\hat{A}|\Psi\rangle = \langle\Psi|\hat{A}|\Psi\rangle. \end{aligned} \quad (\text{A11})$$

However, this time the variance  $\sigma_{\hat{A}}^2$  is only upper bounded by the variance  $\sigma_A^2$  of operator  $\hat{A}$ . This follows from,

$$\begin{aligned} &\sum_s Q(s)|A(s)|^2 \\ &= \sum_s \langle\Psi|\pi(s)|\Psi\rangle \frac{\langle\Psi|\hat{A}^\dagger\pi(s)|\Psi\rangle}{\langle\Psi|\pi(s)|\Psi\rangle} \frac{\langle\Psi|\pi(s)\hat{A}|\Psi\rangle}{\langle\Psi|\pi(s)|\Psi\rangle} \\ &= \sum_s \frac{\langle\Psi|\hat{A}^\dagger\pi(s)|\Psi\rangle\langle\Psi|\pi(s)\hat{A}|\Psi\rangle}{\langle\Psi|\pi(s)|\Psi\rangle} \\ &\leq \sum_s \langle\Psi|\hat{A}^\dagger\pi(s)\hat{A}|\Psi\rangle = \langle\Psi| \left( |\hat{A}|^2 \right) |\Psi\rangle. \end{aligned} \quad (\text{A12})$$

Here, the inequality follows from  $\langle x|y\rangle\langle y|x\rangle \leq \langle x|x\rangle\langle y|y\rangle$  with the identifications  $|x\rangle \equiv \pi(s)\hat{A}|\Psi\rangle$  and  $|y\rangle \equiv \pi(s)|\Psi\rangle$ .

<sup>1</sup> S.R. White, Phys. Rev. Lett. **69**, 2863 (1992).

<sup>2</sup> S.R. White, Phys. Rev. B, **48**, 10345 (1993).

<sup>3</sup> M. Fannes, B. Nachtergaele, and R. F. Werner, Commun. Math. Phys. **144**, 443 (1992).

<sup>4</sup> S. Ostlund and S. Rommer, Phys. Rev. Lett. **75**, 3537 (1995).

<sup>5</sup> G. Vidal, Phys. Rev. Lett., **91**, 147902 (2003)

<sup>6</sup> D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, Quant. Inf. Comput. **7**, 401 (2007).

<sup>7</sup> U. Schollwoeck, Rev. Mod. Phys., **77**, 259 (2005).

<sup>8</sup> U. Schollwoeck, Ann. of Phys. **326**, 96 (2011).

<sup>9</sup> G. Vidal, Phys. Rev. Lett., **99**, 220405 (2007); G. Vidal, Phys. Rev. Lett., **101**, 110501 (2008).

<sup>10</sup> G. Vidal, in *Understanding Quantum Phase Transitions*, edited by L. D. Carr (Taylor & Francis, Boca Raton, 2010), arXiv:0912.1651v2.

<sup>11</sup> F. Verstraete, and J. I. Cirac, arXiv:cond-mat/0407066v1 (2004).

- <sup>12</sup> G. Sierra and M.A. Martin-Delgado, arXiv:cond-mat/9811170v3 (1998).
- <sup>13</sup> T. Nishino and K. Okunishi, J. Phys. Soc. Jpn., **67**, 3066, 1998.
- <sup>14</sup> V. Murg, F. Verstraete, and J. I. Cirac, Phys. Rev. A, **75**, 033605 (2007).
- <sup>15</sup> J. Jordan, R. Orus, G. Vidal, F. Verstraete, and J. I. Cirac, Phys. Rev. Lett., **101**, 250602 (2008).
- <sup>16</sup> Z.-C. Gu, M. Levin, and X.-G. Wen, Phys. Rev. B, **78**, 205116 (2008).
- <sup>17</sup> H. C. Jiang, Z. Y. Weng, and T. Xiang, Phys. Rev. Lett., **101**, 090603 (2008).
- <sup>18</sup> G. Evenbly and G. Vidal, Phys. Rev. B, **81**, 235102 (2010).
- <sup>19</sup> G. Evenbly and G. Vidal, New J. Phys., **12**, 025007 (2010).
- <sup>20</sup> L. Cincio, J. Dziarmaga, and M. M. Rams Phys. Rev. Lett., **100**, 240603 (2008).
- <sup>21</sup> G. Evenbly and G. Vidal, Phys. Rev. Lett., **102**, 180406 (2009).
- <sup>22</sup> Exploitation of space symmetries (e.g. translation invariance and scale invariance) can significantly reduce the computational cost of simulations from  $O(L)$  to  $O(\log L)$  or even to a constant (independent of  $L$ ), allowing to reach the thermodynamic limit.
- <sup>23</sup> N. Schuch, M.M. Wolf, F. Verstraete, J.I. Cirac, Phys. Rev. Lett. **100**, 040501 (2008)
- <sup>24</sup> A. W. Sandvik, G. Vidal, Phys. Rev. Lett. **99**, 220602 (2007).
- <sup>25</sup> S. R. White, Phys. Rev. Lett. **102**, 190601 (2009).
- <sup>26</sup> Y. Y. Shi, L.-M. Duan and G. Vidal, Phys. Rev. A, **74**, 022320 (2006).
- <sup>27</sup> L. Tagliacozzo, G. Evenbly, and G. Vidal, Phys. Rev. B **80**, 235127 (2009).
- <sup>28</sup> E. M. Stoudenmire and S. R. White, New J. Phys. **12**, 055026 (2010).
- <sup>29</sup> A. J. Ferris, G. Vidal, e-print arXiv:1201.3975 (2012).
- <sup>30</sup> H. Katsura, I. Maruyama, J. Phys. A: Math. Theor. **43**, 175003 (2010).
- <sup>31</sup> V. Murg, V. E. Korepin, F. Verstraete, arXiv:1201.5636 (2012) and arXiv:1201.5627 (2012).
- <sup>32</sup> L. Cincio, G. Vidal, *in preparation*.
- <sup>33</sup> From an MPS (TTN) for a state  $|\Psi\rangle$ , with given bond dimension  $\chi$ , one can always use the gauge freedom in these tensor networks to obtain a unitary MPS (respectively TTN) for the same state  $|\Psi\rangle$  and with the same bond dimension  $\chi$ , by writing the tensor network in its canonical form<sup>5,26</sup>.
- <sup>34</sup> We have found some tensor networks for which perfect sampling costs one more power of  $\chi$  than a Markov chain sweep, see Ref. 29. Further, for local operators supported on two or more sites, perfect sampling may incur a slightly larger cost than that of a single sweep of Markov chain Monte Carlo. For example, this occurs with a TTN, where perfect sampling of a two-site operator costs  $\chi^4$  instead of  $\chi^3$ , but not with MPS or MERA.
- <sup>35</sup> The transverse Ising model contains a  $\mathbb{Z}_2$  symmetry as the operator  $\prod_i \hat{X}_i$  commutes with the Hamiltonian. For a wave-function chosen from one of the  $\pm 1$  sectors of this operator, after making a projective measurement of all the spins in the  $x$ -basis, one always gets an even (odd) number number of spin downs (actually, lefts). The overlap after flipping a single spin of such a configuration is always zero, and thus a two-site Markov chain update scheme that can preserve parity is required.
- <sup>36</sup> Our Hamiltonian is actually defined in terms of bond operators, leading to effectively half the magnetic field at the ends of the chain. This change does not affect any of the critical properties of the system.
- <sup>37</sup> We assume that the variance  $\langle \Psi | (\hat{\sigma}_{L/2}^z)^2 | \Psi \rangle - (\langle \Psi | \hat{\sigma}_{L/2}^z | \Psi \rangle)^2$ , which is upper bounded by 1, is essentially constant as a function of the system size  $L$ .